RANGE-PRESERVING UNBIASED ESTIMATORS IN THE MULTINOMIAL CASE

by

Wassily Hoeffding

University of North Carolina at Chapel Hill

Summary Consider estimating the value of a real-valued function $f(\underline{p})$, $\underline{p} = (p_0, p_1, \ldots, p_r)$, on the basis of an observation of the random vector $X = (X_0, X_1, \ldots, X_r)$ whose distribution is multinomial $(n, \underline{p})$. It is known that an unbiased estimator exists if and only if f is a polynomial of degree at most n, and in this case the unbiased estimator of $f(\underline{p})$ is unique. However, in general, this estimator has the serious fault that it is not range-preserving; that is, its values may fall outside the range of $f(\underline{p})$. In this paper, a condition on f is derived which is necessary for the unbiased estimator to be range-preserving, and is sufficient when n is large enough.

Key words and phrases: Range-preserving estimator, unbiased estimator, prior range, posterior range, binomial distribution, multinomial distribution.

1. Introduction

For clarity, the results will first be stated and proved for the binomial case. The extension to the multinomial case, which is straightforward except for the more complicated notation, will be dealt with in the last section of the paper.

Let the random variable X have the binomial (n,p) distribution, $0 \le p \le 1$, and consider the unbiased estimation, based on an observation of X, of a real-valued function f(p). It is known that an unbiased estimator of f(p) exists if and only if f is a polynomial of degree at most n, say

$$f(p) = a_0 + a_1 p + \ldots + a_m p^m, \tag{1.1}$$

where $m \le n$. Since

$$p^k = \sum_{x=0}^{n} \frac{x^{(k)}}{n^{(k)}} \binom{n}{x} p^x (1-p)^{n-x}, \quad k=0,1,\ldots,n,$$

where $n^{(k)} = n(n-1)\ldots(n-k+1)$, an unbiased estimator $t_n(X)$ of f(p) in (1.1) is given by

$$t_n(x) = \sum_{k=0}^{m} a_k x^{(k)}/n^{(k)}, \quad x=0,1,\ldots,n; \quad n \ge m . \tag{1.2}$$

By the completeness of the binomial family (Lehman (1959)) this is the only unbiased estimator of f(p).

In general the values of $t_n(x)$ can fall outside the range of f(p). Thus the unbiased estimator of p(1-p) is x(n-x)/n(n-1), and its maximum with respect to x exceeds 1/4 = max p(1-p).

The main results for the binomial case, proved in Sections 3 and 4 respectively, are the following two theorems.

<u>Theorem 1</u> Let f be a non-constant polynomial. In order that a range-

preserving unbiased estimator of $f(p)$, $0 \leq p \leq 1$, exist, it is necessary that either

(a) $f(0) < f(p) < f(1)$ for $0 < p < 1$; $f'(0) > 0$, $f'(1) > 0$,

or

(b) $f(1) < f(p) < f(0)$ for $0 < p < 1$; $f'(0) < 0$, $f'(1) < 0$.

The necessary condition of Theorem 1 is not sufficient for the unbiased estimator $t_n(x)$ of $f(p)$ to be range-preserving whenever $n \geq \deg(f)$. Thus the polynomial $f(p) = (p-c)^3$, $0 < c < 1$, satisfies condition (a), but for $c > 1/2$, $t_3(1) = c^2(1-c) > (1-c)^3 = f(1)$. However, we have the following result.

Theorem 2 Let f be a polynomial which satisfies condition (a) or (b) of Theorem 1. Then there exists a number $N(f) \geq \deg(f)$ such that the unbiased estimator $t_n(x)$ of $f(p)$ is range-preserving for $n \geq N(f)$.

The needed definitions are given in Section 2.

2. Definitions

Let $\mathcal{P}$ be a family of probability distributions on a measurable space $(\mathcal{X}, \mathcal{A})$ and consider estimating the value of a function $\theta(P)$ defined for $P \in \mathcal{P}$ which takes values in the k-dimensional euclidean space $\mathbb{R}^k$. An estimator $t(x)$ of $\theta(P)$ is a measurable function from $\mathcal{X}$ to $\mathbb{R}^k$. In accordance with [1], the set $\Theta = \{\theta(P): P \in \mathcal{P}\}$ is called the prior range of $\theta(P)$. Informally, the posterior range of $\theta(P)$ given the observation x from $\mathcal{X}$, denoted Θ_x, is the least set in which $\theta(P)$ is known to lie when the value x has been observed. A general definition of the posterior range is given in [1]. In the present binomial case, $\mathcal{P} = \mathcal{P}_n$ is the family of the binomial (n,p) distributions P_p, $0 \leq p \leq 1$, with n fixed. Let $f(p) = \theta(P_p)$. The prior range of $f(p)$ is $\Theta = \{f(p): 0 \leq p \leq 1\}$. The posterior range Θ_x is given by

$$\Theta_x = \{f(p): 0 < p < 1\} \quad \text{if } x = 1, 2, \ldots, n-1, \tag{2.1}$$

$$\Theta_0 = \{f(p): 0 \leq p < 1\}, \quad \Theta_n = \{f(p): 0 < p \leq 1\}.$$

An estimator $t(x)$ of $\theta(P)$ is said to be range-preserving if its values are confined to the posterior range of $\theta(P)$: $t(x) \in \Theta_x$, $x \in X$.

3. Proof of Theorem 1

We need the following result ([1], Proposition 3.2), stated in the setting described in Section 2.

Proposition A Suppose there are two members P_0 and P_1 of $\mathcal{P}$ such that

a) the convex hull of Θ has a supporting hyperplane H at the point $\theta(P_0)$, and

b) P_1 is absolutely continuous with respect to P_0, and $\theta(P_1) \notin H$.

Then no unbiased estimator of $\theta(P)$ is range-preserving.

In our problem the prior range Θ of $f(p)$ is the closed interval $[\min\{f(p); 0\le p\le 1\}, \max\{f(p): 0\le p\le 1\}]$. Suppose $f(p)$ attains its maximum in $[0,1]$ at a point $p_0 \neq 0,1$. Let p_1 be any point in the open interval $(0,1)$ such that $f(p_1) \neq f(p_0)$. (Since f is a non-constant polynomial, p_1 can be so chosen.) The binomial distributions (n,p_0) and (n,p_1) are absolutely continuous with respect to each other. Proposition A (with H being the point $f(p_0)$) implies that the unbiased estimator of $f(p)$ is not range-preserving. Hence, in order that it be range-preserving, it is necessary that $f(p)<f(0)$ or $<f(1)$ for $0<p<1$. Similarly, we must have $f(p)>f(0)$ or $>f(1)$ for $0<p<1$. This proves the necessity of the first conditions in (a) or (b).

To complete the proof, first assume that $f(0)<f(p)<f(1)$ for $0<p<1$. It is sufficient to show that the unbiased estimator $t_n(x)$ of $f(p)$ is range-preserving only if $f'(0)>0$ and $f'(1)>0$.

According to the definitions in Section 2, for $t_n(x)$ to be range-preserving it is necessary that $t_n(1)>f(0)$ and $t_n(n-1)<f(1)$. By (1.2) and (1.1),

$$t_n(1) = a_0 + a_1 n^{-1} = f(0) + f'(0)n^{-1},$$

$$t_n(n-1) = \sum_0^m a_k - \sum_0^m k a_k n^{-1} = f(1) - f'(1)n^{-1},$$

and the stated conditions follow.

For case (b) the completion of the proof is similar.

Note that the necessity of the first condition in (a) or (b) of Theorem 1 has been deduced from Proposition A and deals only with the prior range of $f(p)$. In contrast, the conditions involving the signs of the derivatives of f at 0 and 1 have been obtained from the requirement that the values of the estimator must be confined to the posterior range of $f(p)$.

4. Proof of Theorem 2

We shall assume that the polynomial

(4.1) $$f(p) = a_0 + a_1 p + \ldots + a_m p^m \qquad (m \geq 1)$$

satisfies condition (a) of Theorem 1:

(4.2) $$f(0) < f(p) < f(1) \text{ for } 0<p<1; \; f'(0)>0, \; f'(1)>0.$$

The case (b) can be reduced to case (a) by a simple change of notation.

We must show that there exists an integer $N(f) \geq \deg(f)$ such that for $n \geq N(f)$ the unbiased estimator

(4.3) $$t_n(x) = \sum_{k=0}^m a_k x^{(k)}/n^{(k)}$$

is range-preserving. The latter is true if and only if

(4.4) $$f(0) < t_n(x) < f(1), \quad x=1,2,\ldots,n-1,$$

(4.5) $\quad f(0) \le t_n(0) < f(1), \quad f(0) < t_n(n) \le f(1).$

By (4.3), $t_n(0)=a_0=f(0)$ and $t_n(n)=\sum_0^m a_k = f(1)$, so that conditions (4.5) are satisfied. It remains to show that there is an N(f) such that (4.4) holds for $n \ge N(f)$.

We shall first show that there is a positive $\varepsilon=\varepsilon(f)$ such that

(4.6) $\quad f(0) < t_n(x) < f(1)$ if $0 < x/n < \varepsilon$ or $1-\varepsilon < x/n < 1$.

The proof will be completed by showing that for every $\varepsilon>0$ there is a number $N=N(f,\varepsilon)$ such that for $n \ge N$

(4.7) $\quad f(0) < t_n(x) < f(1)$ if $\varepsilon \le x/n \le 1-\varepsilon$.

Since $x^{(k)}/n^{(k)} \le (x/n)^k$, we have $t_n(x)-f(0) = \sum_1^m a_k\, x^{(k)}/n^{(k)} \le \sum_1^m a_k^+ (x/n)^k$, where $a^+ = \max(a,0)$. Choose $\varepsilon_1 = \varepsilon_1(f) > 0$ so that

$$\sum_1^m a_k^+ \varepsilon_1^k < f(1) - f(0).$$

Then

$$t_n(x) < f(1) \qquad \text{if } x/n \le \varepsilon_1.$$

Also, with $a^- = \max(-a,0)$,

$$t_n(x)-f(0) \ge a_1\, x/n - \sum_2^m a_k^-(x/n)^k$$
$$= x/n(a_1 - \sum_2^m a_k^-(x/n)^{k-1}),$$

where $a_1=f'(0)>0$. Choose $\varepsilon_2 = \varepsilon_2(f) > 0$ so that

$$\sum_2^m a_k^- \varepsilon_2^{k-1} < a_1.$$

Then

$$t_n(x) > f(0) \qquad \text{if } 0 < x/n \le \varepsilon_2.$$

In a similar way one shows that the positive number $\varepsilon_3(f)$ can be so chosen that

$$f(0)<t_n(x)<f(1) \quad \text{if } 0 < (n-x)/n \le \varepsilon_3(f).$$

Setting $\varepsilon(f) = \min_{i=1,2,3}\varepsilon_i(f)$, we have shown that for every polynomial f which satisfies condition (4.2) there is a positive $\varepsilon=\varepsilon(f)$ such that (4.6) holds.

Let $\varepsilon \in (0,\frac{1}{2})$. We must show the existence of a number $N=N(f,\varepsilon)$ such that (4.7) holds for $n\ge N$.

We have

$$f(\tfrac{x}{n}) - t_n(x) = \sum_2^m a_k\left(\frac{x^k}{n^k} - \frac{x^{(k)}}{n^{(k)}}\right).$$

It can be shown by induction on k that

$$(4.8) \qquad 0 \le \frac{x^k}{n^k} - \frac{x^{(k)}}{n^{(k)}} \le \binom{k}{2}\frac{x(n-x)}{n^2(n-1)}, \quad k=2,3,\ldots,n.$$

Since $x(n-x)/n^2 \le 1/4$, we obtain

$$|f(\tfrac{x}{n}) - t_n(x)| \le \frac{1}{4(n-1)}\sum_2^m \binom{k}{2}|a_k|.$$

For a fixed $\varepsilon \in (0,\frac{1}{2})$ define $\overline{f}_\varepsilon = \max\{f(p):\varepsilon\le f(p)\le 1-\varepsilon\}$, $\underline{f}_\varepsilon = \min\{f(p): \varepsilon\le f(p)\le 1-\varepsilon\}$. Note that $\underline{f}_\varepsilon>f(0)$, $\overline{f}_\varepsilon<f(1)$. Hence we can choose a positive $\eta = \eta(f,\varepsilon)$ such that

$$\underline{f}_\varepsilon - \eta > f(0), \qquad \overline{f}_\varepsilon + \eta < f(1).$$

Now choose $N = N(f,\varepsilon)$ so that

$$\frac{1}{4(N-1)}\sum_2^m \binom{k}{2}|a_k| \le \eta.$$

Then if $\varepsilon \le x/n \le 1-\varepsilon$ and $n\ge N$,

$$\underline{f}_\varepsilon - \eta \le t_n(x) \le \overline{f}_\varepsilon + \eta$$

and therefore $f(0) < t_n(x) < f(1)$. This proves the statement around (4.7). //

5. The multinomial case

Let the random vector $\underline{X} = (X_0, X_1, \ldots, X_r)$ take values in the set $\mathcal{X}$ of points $\underline{x} = (x_0, x_1, \ldots, x_r)$ with nonnegative integer-valued coordinates whose sum is n, and let $\underline{X}$ have the multinomial $(n, \underline{p})$ distribution:

$$(5.1) \qquad \Pr\{\underline{X} = \underline{x}\} = \frac{n!}{\Pi_0^r x_i!} \Pi_0^r p_i^{x_i}, \; \underline{x} \in \mathcal{X}, \; \underline{p} \in \underline{\Sigma}_r,$$

$$(5.2) \qquad \underline{\Sigma}_r = \{\underline{p} = (p_0, \ldots, p_r) : p_0 \geq 0, \ldots, p_r \geq 0, \; \textstyle\sum_0^r p_i = 1\}.$$

It will be assumed that the unknown parameter vector $\underline{p}$ varies over the entire simplex $\underline{\Sigma}_r$, the dimension r as well as the sample size n being fixed.

Consider the unbiased estimation, based on an observation of $\underline{X}$, of the value of a real-valued function $f(\underline{p})$, $\underline{p} \in \underline{\Sigma}_r$. As in the binomial case, an unbiased estimator of $f(\underline{p})$ exists if and only if f is a polynomial of degree at most n, and then the unbiased estimator is unique. The latter follows from the completeness of the multinomial family (Lehmann [2], p 132).

We shall be concerned with the local behavior of f at the vertices $\underline{\delta}_0, \underline{\delta}_1, \ldots, \underline{\delta}_r$ of $\underline{\Sigma}_r$, where $\underline{\delta}_i = (\delta_{i0}, \delta_{i1}, \ldots, \delta_{ir})$, $\delta_{ij} = 1$ or 0 according as $i = j$ or $i \neq j$. For $i = 0, 1, \ldots, r$, let $\underline{p}^i = (p_0, \ldots, p_{i-1}, p_{i+1}, \ldots, p_r)$ be the vector $\underline{p}$ with the component p_i deleted. Denote by $f_i(\underline{p}^i)$ the polynomial $f(\underline{p})$ with p_i replaced by $1 - \sum_{j \neq i} p_j$ and write

$$(5.3) \qquad f_i(\underline{p}^i) = \textstyle\sum_m a^i_{j_i \ldots j_r} \, p_0^{j_1} \ldots p_{i-1}^{j_i} \, p_{i+1}^{j_{i+1}} \ldots p_r^{j_r}, \; i = 0, 1, \ldots, r,$$

where the sum $\sum_m$ is extended over the nonnegative integers $j_1, \ldots, j_r$ whose sum is $\leq m = \deg(f)$. Thus (5.3) is the Taylor expansion of $f_i(\underline{p}^i)$ at $\underline{p}^i = (0, \ldots, 0)$. It may be thought of as the expansion of $f(\underline{p})$ at the vertex $\underline{\delta}_i$ of $\underline{\Sigma}$.

Note that $a^i_{0\ldots0} = f_i(0,\ldots,0) = f(\underline{\delta}_i)$, and $a^i_{10\ldots0},\ldots,a^i_{0\ldots01}$ are the first order partial derivatives of f_i at the point $\underline{p}^i = (0,\ldots,0)$.

Due to the identity

$$(5.4)\qquad \Pi_0^r\, p_i^{j_i} = \sum_{\underline{x}\in X} \frac{\Pi_0^r x_i^{(j_i)}}{n^{(\sum_0^r j_i)}} \; \frac{n!}{\Pi_0^r x_i!} \; \Pi_0^r\, p_i^{x_i},$$

it follows from (5.3) that for $n \geq \deg(f)$ the unbiased estimator $t_n(\underline{x})$ of $f(\underline{p})$, $\underline{p} \in \underline{\Sigma}_r$ can be written in each of the forms

$$(5.5)\qquad t_{ni}(\underline{X}^i) = \sum_m a^i_{j_1\ldots j_r} X_0^{(j_1)} \ldots X_{i-1}^{(j_i)} X_{i+1}^{(j_{i+1})} \ldots X_r^{(j_r)} / n^{(\sum_1^r j_i)}, \qquad i=0,1,\ldots,r,$$

where $\underline{X}^i = (X_0,\ldots,X_{i-1},X_{i+1},\ldots,X_r)$.

The following two theorems are extensions of Theorems 1 and 2 from the binomial to the multinomial case.

Theorem 3 Let the random vector $\underline{X}$ have a multinomial $(n,\underline{p})$ distribution, $\underline{p} \in \underline{\Sigma}_r$, and let $f(\underline{p})\colon \underline{\Sigma}_r \to \underline{R}^1$, be a non-constant polynomial, $\deg(f) \leq n$. In order that the unbiased estimator $t_n(\underline{x})$ of $f(\underline{p})$ be range-preserving it is necessary that

(i) both the maximum and the minimum of $f(\underline{p})$ in $\underline{\Sigma}_r$ are attained only at one (or more) of the vertices of $\underline{\Sigma}_r$;

(ii) if the minimum of $f(\underline{p})$ in $\underline{\Sigma}_r$ is attained at the vertex $\underline{\delta}_i$, the first partial derivatives of $f_i(\underline{p}^i)$ at $\underline{p}^i = (0,\ldots,0)$ are all positive;

(iii) if the maximum of $f(\underline{p})$ in $\underline{\Sigma}_r$ is attained at the vertex δ_j, the first partial derivatives of $f_j(\underline{p}^j)$ at $\underline{p}^j = (0,\ldots,0)$ are all negative.

<u>Theorem 4</u> Let f be a polynomial which satisfies the conditions of Theorem 3. Then there exists a number N(f) such that the unbiased estimator t_n of $f(\underline{p})$ is range-preserving for $n \geq N(f)$.

<u>Proof of Theorem 3</u>: The necessity of condition (i) follows from Proposition A of Section 3.

Suppose that the minimum of $f(\underline{p})$ in $\underline{\Sigma}$ is attained at the vertex $\underline{\delta}_i$. Then, for $t_n(\underline{x})$ to be confined to the posterior range of f, it is necessary that $t_n(\underline{x}) > f(\delta_i)$ if $x_i = n-1$ (hence $x_k = 1$ for some $k \neq i$ and $x_\ell = 0$ for $\ell \neq i,k$). By (5.5) this implies condition (ii).

The necessity of condition (iii) follows in a similar way.

<u>Proof of Theorem 4</u>: As in the proof of Theorem 2, it is first shown that if $\underline{x}/n$ is in a sufficiently small ε-neighborhood of one of the vertices $\underline{\delta}_i$ at which $f(\underline{p})$ attains its maximum or its minimum, but $\overline{x}/n \neq \delta_i$, then $\min f(\underline{p}) < t_n(\underline{x}) < \max f(\underline{p})$. Here conditions (ii) and (iii) of Theorem 3 are used. Then it is shown that if $\underline{x}/n \in \underline{\Sigma}_r$, $2 \leq \sum_1^r j_i \leq m$, we have

$$\left| \Pi_1^r \left(\frac{x_i}{n}\right)^{j_i} - \Pi_1^r \, x_i^{(j_i)} \Big/ n^{(\sum_1^r j_i)} \right| \leq C_{r,m} n^{-1}, \tag{5.6}$$

where $C_{r,m}$ depends only on r and m. For r = 1, (5.6) follows from (4.8); in the general case it can be proved by induction on r. Since

$$f\left(\frac{x}{n}\right) - t_n(\underline{x}) = \sum_{2 \leq \sum_1^r j_i \leq m} \overset{\circ}{a}_{i,\ldots,i_m} \left\{ \Pi_1^r \left(\frac{x_i}{n}\right)^{j_i} - \Pi_1^r x_i^{(j_i)} \Big/ n^{(\sum_1^r j_i)} \right\},$$

(5.6) implies

$$\left| f\left(\frac{x}{n}\right) - t_n(\underline{x}) \right| \leq C(f) n^{-1}, \tag{5.7}$$

where C(f) depends only on f.

Using (5.7), the proof is completed by an argument similar to that used in the proof of Theorem 2.

References

[1] Wassily Hoeffding (1983) Unbiased range-preserving estimators. A Festschrift for Erich L Lehmann, Wadsworth, Belmont, pp 249-260.

[2] E. L. Lehmann (1959) Testing Statistical Hypotheses, Wiley, New York.